Regression Models Course Project

Joy SN

January 5, 2019

How is MPG affected by Transmission Type (Automatic and Manual)?

Promblem Summary

Study Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, explored the relationship between a set of variables and miles per gallon (MPG) (outcome). Two questions of importance

“Is an automatic or manual transmission better for MPG”
“Quantify the MPG difference between automatic and manual transmissions”

Exploratory Data Analysis

Dataset: Motor Trend Car Road Tests https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html.

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs Engine (0 = V-shaped, 1 = straight) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

The *figure 1*, we see that correlation between different continuos variables. Per the *Figure 1*, MPG is not so strongly related to drat and qsec

Statistical Inference

Cars with “manual”" transmission have with mean MPG of 24.4 miles/gallon Cars with “automatic” tranmission have mean MPG of 17.1 miles/gallon

```
##
##  Welch Two Sample t-test
##
## data:  data[data$am == 1, ]$mpg and data[data$am == 0, ]$mpg
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

Cars with manual transmission Average MPG is 24.39231 miles/gallon compared to 17.14737 MPG in automatic transmission.
95% CI: 3.2 - 11.3 ,p = 0.0014).

Model Genration

We start with a model with all the columns as perdictors and based on different P-values we remove each columns

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

Since the P value of cyl is very high, we remove it and build another model minus cyl

Then we see that the P value of 'vs' is quite high, so we now build another one without 'VS' and so on we remove 'carb', 'gear', 'drat','disp' and 'hp' because of their high P value

As the p-value of all the remaining predictors are smaller than 0.05, we can stop. Finally we are leftt with 3 variables - wt, qsec and am
- am Transmission (0 = automatic, 1 = manual)
- mpg Miles/(US) gallon
- wt Weight (1000 lbs)
- qsec 1/4 mile time

```
##       Estimate Std. Error   t value     Pr(>|t|)
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## am    4.299519  1.0241147  4.198279 2.329423e-04
```

As the two-sided p-value for the coefficient of `am` is 2.329423110^{-4}, much smaller than 0.05, we can reject the hypothesis H_0.

With 95% confidence, we estimate that a the change from automatic to manual transmission results in a 2.2049694 to 6.3940689 increase in miles per gallon for the cars. Looks like the manual transmission is better than automatic transmission for mpg. Is it causal?

But there is a caveat. Per *Figure 2* , which is a pair plot among different predictors we see a relation ship between transmission (am) and weight. Also in *Figure 3*, which is a box plot of, we see there is a strong relation between weight and Transmission

Final Modelling

We start with Sequential Testing

1.Model 1 - MPG ~ wt (Weight)

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

R Squared :- 0.74
SD :- 3

2. Model 2 - MPG ~ wt (Weight) + am (Transmission Type) Next we add column AM (Transmission type)

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

R Squared :- 0.74
SD :- 3
We see that the addition of the variable "am"(transmission type) to Model 1 in Model 2 does not change the values much. Thus it is not addition any additional value to explain MPG.

```
anv1<-anova(fit1, fit2)
```

Using anova, we get
RSS Values 278.3219375, 278.3196972 P Value 0.9879146

The effect that "transmission on"MPG"" is mostly due to its relation to weight which was described above (manual cars are significantly lighter than larger cars)

But as seen in *Figure 4*, there appears to be an interaction between weight and type of transmission to explain MPG. There are different different slopes for weight for each type of transmission, with manual transmission showing a greater decrease in MPG as the weight of the car increases.

3. Model 3 - mpg ~ wt * am

This model includes the interaction (weight*transmission)

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 31.416055  3.0201093 10.402291 4.001043e-11
## wt          -3.785908  0.7856478 -4.818836 4.551182e-05
## am          14.878423  4.2640422  3.489277 1.621034e-03
## wt:am       -5.298360  1.4446993 -3.667449 1.017148e-03

## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + am + wt:am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 278.32
## 2     28 188.01  2    90.314 6.7253 0.004119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R Squared :- 0.82
RSS Values 278.3219375, 188.0076659
In this final model all coefficients are significant
ANOVA shows that there is some inclusion of wt * am term and it reduces RSS.

For “manual” engine cars, for every unit increase in weight(wt) , there is an average drop of MPG = 9.08 miles/gallon For “automatic” cars, this fall in MPG is 4 miles/gallon.

Figure 5 shows the residual and other plots for this final fit. We note that the residuals show no obvious pattern, so it is reasonable to try to fit a linear model to the data. Now with all the previous analysis, we can conclude that our linear model is a resonable fit.

Conclusion/Executive Summary

MPG of Manual engine cars are on an average less than that of automatic engine tarnsmission type cares. This is also due to the fact that “manual” type cars are lighter than “automatic” engine cars. But, for “manual” cars, MPG w.r.t to weight is effected higher (9.08 miles/gallon) than that of “automatic” cars (4 miles/gallon)

Appendix A

Fig 1 - Shows the correlation of continuous data

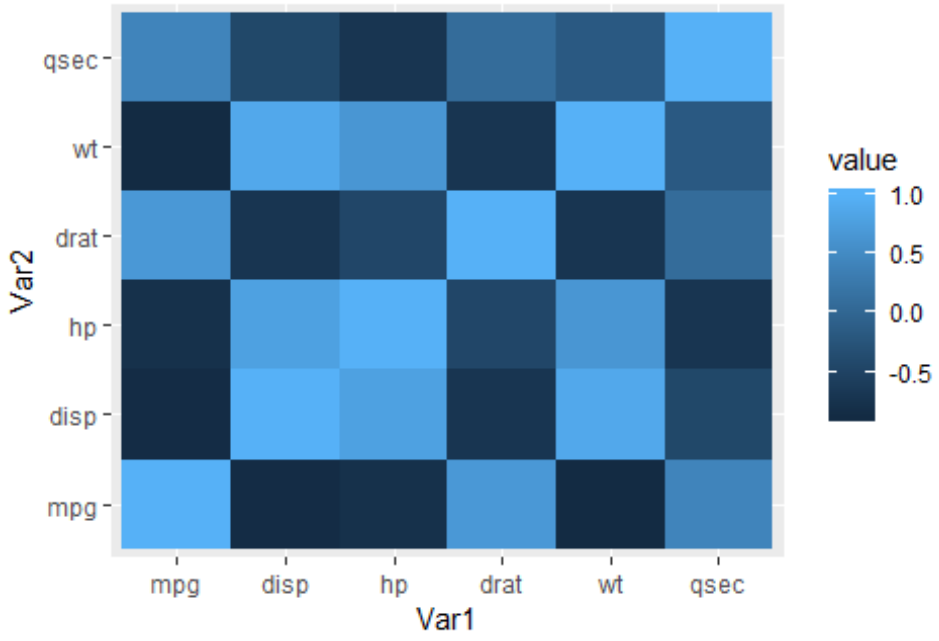

Fig 2 - Pair Plot among mpg, wt,qsec and am

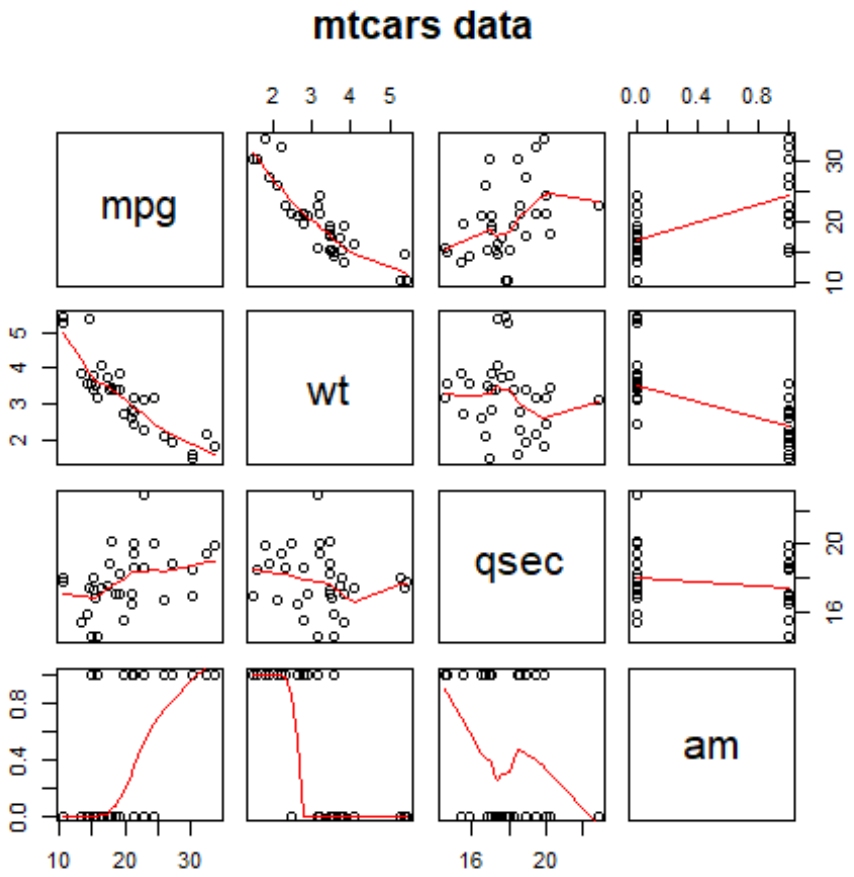

Fig 3 Wights w.r.t MPG for both types of Transmission Type

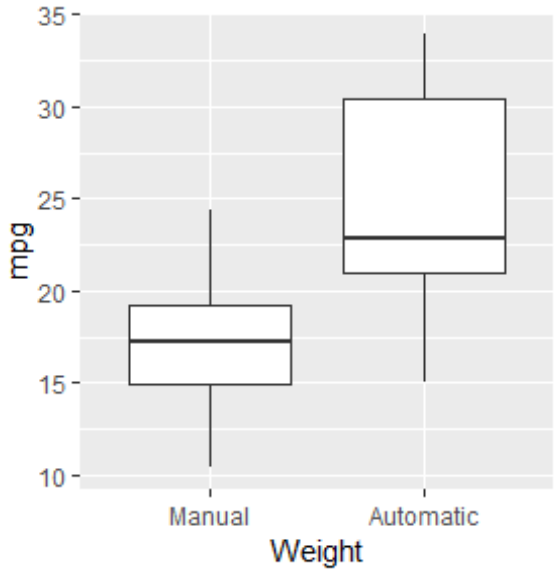

Fig 4 - Shows the Mean wights of both transmission type w.r.t MPG

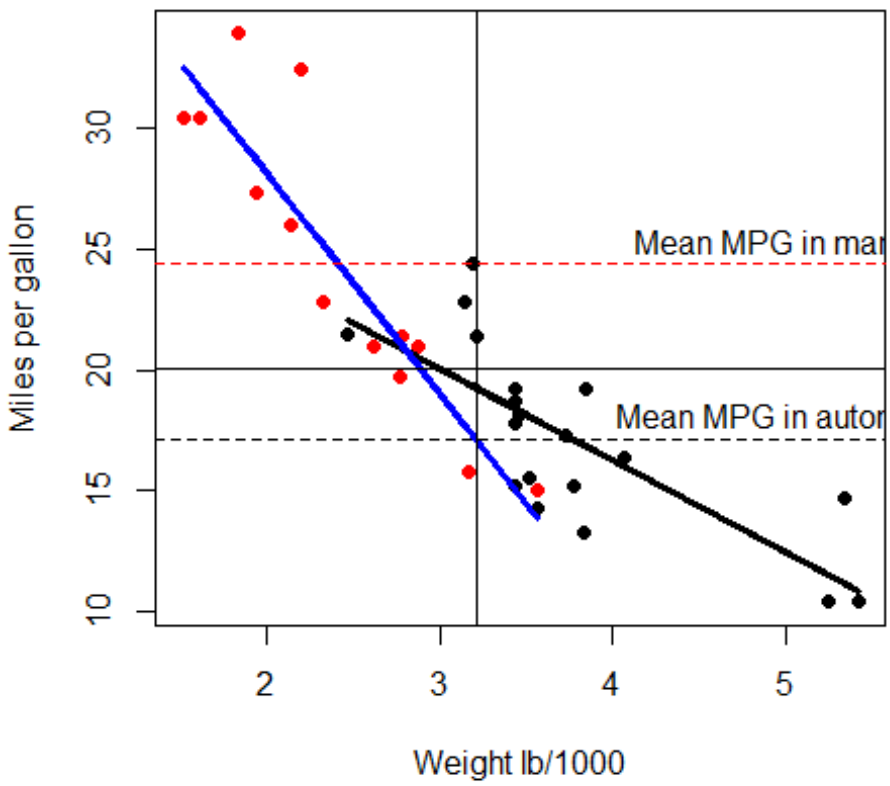

Fig 5 - Residual and other variations